Manuscript published in Journal of Proteomics, [Epub ahead of print], 2019

High sensitivity proteomics of prostate cancer tissue microarrays to discriminate between healthy and cancerous tissue

Lilla Turiák[1], Oliver Ozohanics[1], Gábor Tóth[1,2], András Ács[1,3], Ágnes Révész[1], Károly Vékey[1], András Telekes[4,5], László Drahos[1]*

[1]MS Proteomics Research Group, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar Tudósok körútja 2., H-1117, Budapest, Hungary

[2]Budapest University of Technology and Economics, Faculty of Chemical Technology and Biotechnology, H-1111, Budapest, Műegyetem rkp. 3., Hungary

[3]Semmelweis University, Ph.D. School of Pharmaceutical Sciences, Üllői út 26, H-1085, Budapest, Hungary

[4] Div. Sect. of Geriatrics, 2nd Department of Internal Medicine, Semmelweis University, H-1115, Budapest, Halmi utca 20-22, Hungary

[5] Dept. of Oncology, St Lazarus County Hospital, H-3100, Salgótarján, Füleki út 54-56, Hungary

Corresponding author:

Lilla Turiák, Ph.D.

Research Centre for Natural Sciences, Hungarian Academy of Sciences,

Magyar Tudósok körútja 2., H-1117, Budapest, Hungary

Email: turiak.lilla@ttk.mta.hu

Tel: +36 1 382 6516

Abstract

Biopsies, in the form of tissue microarrays (TMAs) were studied to identify anomalies indicative of prostate cancer at the proteome level. TMAs offer a valuable source of well-characterized biological material. However, because of the small tissue sample size method development was essential to provide the sensitivity and reliability necessary for the analysis. Surface digestion of TMA cores was followed by peptide extraction and shotgun proteomics analysis. About 5 times better sensitivity was achieved by the optimized surface digestion compared to bulk digestion of the same TMA spot and it allowed the identification of over 500 proteins from individual prostate TMA cores. Label-free quantitation showed that biological variability among all samples was about 3 times larger than the technical reproducibility. We have identified 189 proteins which showed statistically significant changes (t-test p-value < 0.05) in abundance between healthy and cancerous tissue samples. The proteomic profile changed according to cancer grade, but did not show a correlation with cancer stage. Results of this pilot study were further evaluated using bioinformatics tools, identifying various protein pathways affected by prostate cancer progression indicating the usefulness of studying TMA cores to identify quantitative changes in tissue proteomics.

Significance

Detailed proteomics analysis of TMAs presents a good alternative for tissue analysis. Here we present a novel method, based on tissue surface digestion and nano-LC-MS measurements, which is capable of identifying and quantifying over 500 proteins from a 1.5 mm diameter tissue section. We compared healthy and cancerous prostate tissues samples, and tissues with various grades and stages of cancer. Tissue proteomics clearly distinguished healthy and cancerous samples, furthermore the results correlated well with cancer grade, but not with cancer stage. Over 100

proteins showed statistically significant abundance changes (t-test p-value < 0.05) between various groups. This was sufficient for a meaningful bioinformatics evaluation; showing e.g. increased abundance of proteins in cancer in the KEGG ribosome pathway, GO mRNA splicing via spliceosome, and chromatin assembly biological processes. The results highlight the feasibility of the developed method for future large-scale tissue proteomics studies using commercially available TMAs.

Keywords: prostate cancer, tissue microarray, mass spectrometry, proteomics, surface digestion, label-free quantitation

1. Introduction

Prostate cancer is among the top three most common types of cancer in men [1]. Tissue biopsies [2, 3] are frequent objects of mass spectrometry (MS) based research and could provide useful insights into understanding biochemical mechanisms of diseases, such as cancer. Detecting anomalies at the molecular level is essential for identifying potential drug targets and for developing new treatment options. Tissue microarrays (TMAs) are a pathologically well-characterized source of biopsies fixed on a microscope slide and arranged in an array format [4]. Cancer stage is known for these TMA specimens and cancer grade is also determined by pathologists, just like in the case of conventional biopsies. Grading describes the morphology of the cells, while staging depicts the extent of spreading of cancer. TMAs are also characterized by

immunohistochemistry [5, 6]. TMAs can be studied by two different MS approaches to acquire information on molecular composition; either by MALDI imaging or by LC-MS/MS. Proteomics studies of various cancer TMA cores using MALDI imaging mass spectrometry have already been reported [7, 8]. These type of analyses generally do not aim to identify specific proteins altered in cancer but detect *m/z* signals (molecular features) that change with different cancer phenotypes and with spatial location. LC-MS/MS analysis is usually another option to study TMA cores [9]. This can be achieved by scraping off TMA cores and performing shotgun mass spectrometry analysis. However, these studies are seriously hampered by the limited amount of sample available (usually 1.5 mm diameter, 5 μm thick cores). To improve sensitivity, several cores are often scraped off and pooled, but these types of studies are better performed on bulk tissue sections, rather than TMAs.

MS-based proteomics on prostate cancer tissues has been reviewed extensively [10-13]. Experimental methodologies include 2DE followed by MALDI-TOF-MS/MS [14]; laser capture micro-dissection and LC-MS/MS [15]; and SILAC-based quantitative LC-MS/MS [16]. Considering the small amount of tissue samples obtained for pathological diagnosis it is important to develop proper sample handling to obtain reliable proteomics data using these specimen. The aim of this preliminary study was to develop and describe a sufficiently sensitive and reliable LC-MS/MS method for label-free quantitative analysis of individual TMA cores. Digesting very low amounts of tissue (like scraping off TMA cores) leads to significant sample loss, so we used surface tissue digestion to improve sensitivity. A previously described tissue surface tissue digestion protocol [17] was adapted to individual prostate cancer TMA cores. Following surface digestion, the resulting peptides were extracted from the surface and analyzed by LC-MS/MS. TMA cores of the same patient as well as patients belonging to the same pathological grade were

tested. Label-free quantitative proteomics was used to detect proteomic changes occurring among prostate cancer samples.

2. Materials and methods

Unless it is stated otherwise, reagents and consumables were purchased from Sigma-Aldrich (Sigma-Aldrich Kft., Budapest, Hungary).

2.1. Surface digestion of TMA cores

Human formalin fixed paraffin embedded (FFPE) prostate cancer tissue microarray slides (T191a and T196) were purchased from US Biomax, Inc. (Derwood, MD, USA). Core diameters are 1.5 mm and tissue thickness is 5 µm. Tissue dewaxing and antigen retrieval of the slides was performed according to a previous protocol [17]. Altogether, the two slides contain 12 different cases (4 cores/each case) corresponding to Normal (n=2), Grade 1 (n=1), Grade 2 (n=4), Grade 2-3 (n=2) and Grade 3 (n=3) samples. Therefore, in total 48 TMA cores were analyzed (Table 1).

Tryptic digestion on the surface of the TMA cores was based on a previously described protocol [17]. Briefly, the proteins in the TMA cores were denaturated and alkylated by 1 µL solution containing 0.1% RapiGest SF (Waters, Milford, MA) + 5 mM DTT + 10% glycerol. The tissue slide was incubated in a humidified box at 55 ºC for 20 min. Next, 1 µL solution containing 25 mM ammonium bicarbonate + 10 mM iodoacetamide (IAA) + 10% glycerol was added to each core and incubated in the dark for 20 minutes. 1 µL Trypsin/Lys-C mix (Promega, Madison, WI) enzyme solution (50 ng/ µL Trypsin/Lys-C mix in 50 mM ammonium bicarbonate and 10% glycerol) was added in two cycles, followed by three cycles of 1 µL Trypsin (Promega, Madison, WI) enzyme solution (200 ng/µL Trypsin in 50 mM ammonium bicarbonate and 10% glycerol).

Each cycle consisted of 40 minutes of incubation at 37 ºC. Total enzymatic digestion time was 200 minutes. Resulting peptides were extracted from the surface of the individual TMA cores manually by repeated pipetting using 4 x 1.5 µL 10% acetic acid extraction solvent on each core. Samples were dried in a SpeedVac and desalted using Pierce C18 spin columns (Thermo Fisher Scientific, Waltham, MA).

2.2. In-solution digestion of TMA cores

To assess the efficiency of surface digestion the obtained proteomics results were compared to parallel TMA cores that were scraped off the microscope slide and digested in-solution. Tissue dewaxing and antigen retrieval of the slides was performed as above [17]. After that individual TMA cores were scraped from the surface of the microscope slide using a needle and transferred to Eppendorf tubes containing 50 µL lysis buffer (100 mM Tris pH 7.6 + 4% SDS) and incubated in a block heater at 97 °C for 30 min. The resulting cell lysate was clarified by centrifugation at 16 000 g for 10 minutes. The supernatant was transferred to clean Eppendorf tubes and 9x volume ice-cold ethanol was added. Proteins were precipitated at -20 ºC overnight. Next day, the pellets were washed twice with ice-cold ethanol and dissolved in 10 µL of 8 M urea in 50 mM ammonium bicarbonate. DTT was added at a final concentration of 5 mM and incubated at 37 °C for 30 min. Alkylation was performed in the dark at room temperature for 30 minutes in the presence of 10 mM IAA. Samples were diluted 10-fold with 50 mM ammonium bicarbonate and 1 µL 10 ng/µL or 50 ng/µL Trypsin/Lys-C mix (Promega, Madison, WI) was added and incubated at 37 °C for 80 minutes. Next, 1 µL 40 ng/µL or 200 ng/µL Trypsin (Promega, Madison, WI) was added and the samples were incubated for another 2 hours. Digestion was quenched by the addition of 1 µL formic acid. Desalting was performed on Pierce C18 spin columns (Thermo Fisher Scientific, Waltham, MA).

Note that using “surface” and the conventional “in-solution” digestions discussed above, different enzyme concentrations were used. We compared the efficiency of the two digestion methods using the same enzyme concentrations and incubation times. The same samples (replicate Grade 2 biopsies) were digested both in solution following scraping off the tissue from the surface of the glass slide, and on tissue surface using both low (10 ng/µL Lys-C/Trypsin mix and 40 ng/µL Trypsin solution) and high enzyme concentrations (50 ng/µL Lys-C/Trypsin mix and 200 ng/µL Trypsin solution). Each condition was tested on two technical replicates.

2.3. Mass spectrometry and chromatography analysis

Samples were dissolved in 12 µL solvent (98% water, 2% acetonitrile and 0.1% formic acid) out of which 6 µL was subjected to nanoLC-MS/MS analysis using a Dionex Ultimate 3000 RSLC nanoLC (Dionex, Sunnyvale, CA, USA) coupled to a Bruker Maxis II Q-TOF (Bruker Daltonik GmbH, Bremen, Germany) via CaptiveSpray nanoBooster ionization source. Peptides were separated on an Acquity M-Class BEH130 C18 analytical column (1.7 µm, 75 µm x 250 mm Waters, Milford, MA) using gradient elution (4–50% eluent B in 120 minutes) following trapping on an Acclaim PepMap100 C18 (5 µm, 100 µm x 20 mm, Thermo Fisher Scientific, Waltham, MA) trap column. Solvent A consisted of water + 0.1% formic acid, while Solvent B was acetonitrile + 0.1% formic acid. Spectra were collected using a fix cycle time of 2.5 sec and the following scan speeds: MS spectra were acquired at 3 Hz, while CID was performed on multiply charged precursors at 16 Hz for abundant ions and at 4 Hz for low abundance ones. Internal calibration was performed by infusing sodium formate and data were automatically recalibrated using the Compass Data Analysis software 4.3 (Bruker Daltonik GmbH, Bremen, Germany).

2.4. Protein identification and label-free quantitation

Data were processed by the ProteinScape 3.0 software (Bruker Daltonik GmbH, Bremen, Germany). Proteins were identified by searching against the human Swissprot database (2015_08) using the Mascot search engine version 2.5 (Matrix Science, London, UK). First, the individual LC-MS results were searched by Mascot (7 ppm peptide mass tolerance, 0.05 Da fragment mass tolerance, 2 missed cleavages, carbamidomethylation of cysteines as fixed modification, deamidation (NQ) and oxidation (M) as variable modifications) and proteins were identified using 1% FDR limit. For additional validation and the comparison of the two types of digestion and the different cancer grades the resulting Mascot search files were loaded into Scaffold 3.0 software [18]. The parameters used were the following: 95% probability, minimum 2 peptides identified, additional X! Tandem identification and all the other parameters were set to default parameters. Second, the Mascot search results were merged by ProteinScape for all 48 samples; and proteins were considered identified using strict criteria (ion score criteria was >20 (corresponding to $p<0.01$), and FDR < 1%). Label-free quantitation (LFQ) was then performed using MaxQuant [19] (software version 1.5.3.30), applying its default parameters. MaxQuant analysis searched only for those proteins, which were identified previously by Mascot (this makes false identification less likely). Strict acceptance criteria were used, minimum 2 peptides, less than 1% FDR was set both at the protein and at the peptide level. Each LC-MS/MS run was aligned using the “match between runs” feature (match time window 0.8 minute, alignment time window 15 minutes). Details of the parameters of this procedure are given in the Supplementary material (see also Table S-1 and S-2).

2.5. Statistics and bioinformatics

In most studies proteins showing the most significant abundance changes between two groups of subjects (healthy/cancerous, grade 2-3) are usually selected based either on the degree of fold-change or the statistical significance (p-value of a t-test). A further aspect to consider is protein abundance: selecting very low abundance proteins based on the fold-change can often be misleading. We decided to use a combination of these values using the following formula, which can be considered as a modified p-value:

$$p_{mod} = p/(fmod \times \sqrt[3]{I}) \qquad \text{Eq 1.}$$

Where $\mathbf{p_{mod}}$ is the modified p-value between healthy and cancerous tissue; the smaller the more significant; $\mathbf{f_{mod}}$ is the modified fold-change, and **I** is the abundance of the protein (LFQ value). When the protein abundance is decreased between two samples, the fold-change is a number between zero and unity. In this case, the modified fold-change will be the reciprocal value of the fold change (so that $\mathbf{f_{mod}}$ is always larger than unity). Due to statistical fluctuations, the fold-change of low abundance peaks is often unreasonably high. To limit the influence of such fluctuations, the modified fold-change is limited to maximum 5 (i.e. it will always be between 1 and 5). Therefore the $\mathbf{p_{mod}}$ will be smaller (more significant) under the following conditions: the p-value is small, the modified fold-change and/or the protein abundance are large. The 200 proteins showing the most significant changes based on $\mathbf{p_{mod}}$ were selected for further study. From this group, those showing both statistical significance ($p < 0.02$) and sufficiently high modified fold change (over 2) between healthy and cancerous samples were selected for protein interaction analysis (see Table S-3a). Similarly, proteins with $p < 0.02$ and modified fold change over 1.4 were selected to compare various cancer grades (see Table S-3b).

Protein interaction analysis was performed using the "Search Tool for Recurring Instances of Neighboring Genes" (STRING) [20]. Functional enrichment analysis was performed using the ClueGO plugin [21] of Cytoscape, and the statistical test used for the enrichment was based on the two-sided hypergeometric option with a kappa score of 0.2. The search was performed against the GO-Uniprot databases and the KEGG database. XLStat software (Addinsoft, Paris, France) was used for running non-supervised PCA analysis.

3. Results and Discussion

3.1. Tissue surface digestion of TMA cores

The method for tissue surface digestion is based on a previously described methodology [17] which was adapted for TMA analysis. Instead of Trypsin, we used a mixture of Trypsin and Lys-C in the present protocol, which provides better (more efficient and more complete) digestion. The analysis was performed using nanoUHPLC-MS/MS using a 2-hour-long linear gradient. Details of the protocol are described in the Materials and Methods part (2.1.).

The number of proteins identified strongly depends on the type and the size of the sample. In the present case 386-589 proteins were identified (using Mascot search, validated by Scaffold) from a single TMA core (1.5 mm diameter, 5 µm thick, ca. 10 µg total mass, 1 µg protein content). The number of identified proteins on average was 518. This compares favorably with results on mouse brain and liver tissue sections, where on average 42 proteins were identified using a similar sample size [17].

We compared tissue surface analysis with digestion of the bulk tissue by analyzing TMA cores in duplicates (Fig.1). The same samples (replicate Grade 2 biopsies) were digested both in solution

following scraping off the tissue from the surface of the glass slide and on tissue surface using both low and high enzyme concentrations as described at the end of section 2.2, in two technical replicates. The four groups were compared, shown in a Venn diagram (Fig. 1). The results show that enzyme concentrations have a relatively small influence. Surface digestion provides between 4.0 and 5.3-fold increase in the number of identified proteins compared to bulk tissue analysis in solution, depending on enzyme concentration. Typical contaminant proteins were identified only in scraped-off tissue samples suggesting that the on-surface digestion is less prone to contamination.

The workflow described in section 2.1. was performed on 4 replicate cores of 12 individuals; altogether 48 individual TMA cores were studied. The individuals were divided into Normal (i.e. healthy), Grade 1, Grade 2, Grade 2-3 and Grade 3 groups, based on histological analysis. Results of Normal, Grade 2 and Grade 3 cancer sample analysis are compared in a Venn diagram (Fig. 2). The Figure shows that significantly more proteins (by 242) were identified in cancerous than in healthy tissue.

3.2.Label-free quantitation, analytical reproducibility, and biological variability

The amount of proteins in the various samples has been determined using label-free quantitation. This was performed by the MaxQuant software [19], using normalized LFQ values. Data are shown in Table S-1, while identified peptides are listed in Table S-2.

The similarity of two samples based on respective protein abundances, like spectral similarities, can be compared using correlation coefficients [22]. Here we used Pearson correlation coefficients to compare protein abundances (MaxQuant LFQ values) among various samples. The correlation coefficients of protein abundances among technical replicates (pair-wise) vary significantly, from 0.99 to 0.68, the average being 0.93. Replicates showing the worst correlation coefficients ("outliers") are due to major, non-statistical variations (e.g. damaged TMA core, mistakes in

sample handling, or temporal changes in ESI spray conditions) [23]. In order to improve the reliability of analysis, these outliers should be left out. To do it in a non-biased way, we discarded the most outlying experiment out of the 4 technical replicates in each case, based on the correlation coefficient. The average correlation coefficient among the remaining 3 technical replicates was 0.95 (24% RSD for the 200 most abundant proteins), and there were no major deviations from the average. These technical replicates were averaged, which reduced the technical error (RSD) of the average abundance to 13.8%, and we compared this value to the biological variability discussed below. Note, the standard deviation of the average is equal to the standard deviation of the dataset divided by the square root of the number of replicates. Technical reproducibility includes errors not only in sample preparation and analysis but variability among various tissue-slices as well (different cell assemblies in a given tissue).

The biological variability among all the samples reflects the influence of prostate cancer on cellular proteomics. This was determined as an average correlation coefficient among the 12 different samples (0.85), and as the average RSD of the 200 most abundant protein intensities (46.7%), shown in Table 2. This clearly indicates that the biological variability is much larger than the technical reproducibility (13.8% RSD). The samples studied may be divided into various groups, based on cancer stage or cancer grade (Table 1). Among the 12 samples studied, there were 2 healthy, 5 Stage II, 2 Stage III, and 3 Stage IV cases. The average biological variability within the four groups of patients can be characterized by an average correlation coefficient of 0.92 and an average RSD of 30.5 % (Table 2). This means that the variability within the various groups is smaller than among all the samples but it is still larger than the technical reproducibility. The same samples were grouped together based on cancer grade as well: there were 2 healthy, 1 Grade 1, 4 Grade 2, 3 Grade 3 cases, and 2 of them were assigned as between Grade 2 and 3. The average biological variability within normal, Grade 2 and Grade 3 cancer types was 0.94 based on the

274 average correlation coefficient, and 28.7% based on average RSD of protein abundances (Table
275 2). These results show that grouping samples based on grade results in somewhat more uniform
276 groups than classification based on stage. All of these data are listed in Table 2.

277 The main purpose of the present communication is to develop a new method and evaluate its
278 potential in TMA proteomics for exploring molecular changes occurring in prostate cancer. It is a
279 pilot study since the number of samples analyzed is not large enough to fully describe biological
280 variations. However, the small number of samples (12) were analyzed in several replicates, so it
281 is feasible to separate technical errors from biological differences. As discussed above, technical
282 errors are significantly smaller than the biological variability, corroborating the robustness of the
283 analytical methodology used. This pilot study is a starting point to map biological differences
284 among various patient groups and for designing a future, large-scale, statistically relevant study.

285 **3.3. Principal component analysis**

286 Quantitative proteomics (arrays of protein abundances) was used to compare the studied 12
287 samples using principal component analysis (PCA), and the result is shown in Fig. 3.

288 The two principal components (linear combination of protein abundances) providing the maximum
289 variability among the samples are plotted on the two axes in Fig. 3. PCA is an unsupervised
290 statistical approach, therefore no information on the sample type (i.e. healthy or stage II cancer) is
291 taken into account; it is important to avoid bias. Fig. 3a shows the results where samples are
292 distinguished based on cancer stage, while Fig. 3b shows the same data labelled according to
293 cancer grade. The ovals in Fig. 3 mark the distribution of the various sample types in the plot.

294 It is shown in Fig. 3 that the healthy samples were clearly separated from all the cancerous
295 specimens. However, Fig. 3a suggests that protein distribution does not separate the various cancer

stages from each other. Furthermore, there is no direct progression from Stage II to III and IV but the individual groups are nearly randomly mixed. This suggests that the overall protein distribution (i.e. major biochemical processes) cannot distinguish cancer stages reasonably. Fig. 3b shows the separation of various grade cancer types. Cancerous samples are clearly separated into two groups: Grade 1, Grade 2 and "uncertain" Grade 2-3 samples belong to one group, while Grade 3 cancer specimens constitute another group. The results suggest that the protein fingerprint distinguishes tissue biopsies based on cancer grade. Note, Grade 3 cancer is defined by histology containing poorly differentiated (i.e. far from healthy) cells. The present results show that the changes in protein fingerprint relate fairly well to the differences observed in cell structure, while only slight correlation is depicted with stage which is mainly determined by the spread of cancer or metastasis. Although the overall protein fingerprint does not correlate with cancer stage, the possibility that there might be specific minor proteins (or protein pathways) which may reflect metastasis is not excluded.

Differences among various cancer groups can be studied using correlation coefficients as well. The average correlation coefficients between protein fingerprints (sets of protein abundances) within and between various groups based on cancer grade are shown in Table 3. The first column shows the comparison with healthy subjects: differences in the protein fingerprint between healthy and cancerous groups increase with higher cancer grade. Table 3 also indicates that samples obtained from Grade 2-3 and Grade 3 subjects are highly heterogeneous, as evidenced by the small correlation coefficients both within and between these groups. These findings support conclusions based on PCA analysis.

3.4. Changes in proteins between healthy and cancerous tissue

Results can be further evaluated by looking at changes in the abundances of individual proteins. According to the PCA results, Normal, Grade 2 and Grade 3 subjects form well-separated groups based on the protein fingerprint. Most proteins can be classified into four groups based on changes in their abundance: a) housekeeping proteins, b) proteins which show decreasing abundance in the healthy >>>Grade 2 >>>Grade 3 direction, c) proteins which show increasing abundance in the healthy <<<Grade 2 <<<Grade 3 direction, and d) proteins which have similar abundance in healthy and Grade 3 but show increased abundance in Grade 2 cancer. Selected examples for these four protein groups are shown in Fig. 4, while the relative abundances of all the detected proteins are shown in the Table S-1. Abundant proteins typically show decreasing abundance in cancer. On the other hand, over 100 proteins show statistically significant increase with cancer progression.

Proteins showing the most significant abundance changes between healthy and cancerous subjects were selected based on a t-test, weighted by the fold change and the protein abundance (those with high fold-change and high abundance were selected preferentially). 15% of all proteins were selected for further analysis; this limit corresponds to t-test p-value lower than 0.02 and fold-change larger than 2. We used bioinformatics analysis (STRING bioinformatics program) [20] to identify protein pathways and protein localizations involved in the tumorigenesis of prostate cancer. The result of STRING analysis is indicated in Fig. 5 using proteins listed in Table S-3a. Using highly confident protein interactions only, 5 protein groups can be distinguished in Fig. 5, outlined by the drawn ovals. All proteins present in Group 1 show increased abundance in cancer. These are ribosomal proteins that participate in biological processes such as translational elongation and protein localization to organelle (based on gene ontology) and belong to the KEGG ribosome pathway. This is in agreement with previous studies, showing that various ribosomal proteins are up-regulated in prostate cancer [24, 25]. Ribosomal protein-based cancer signature has been recently proposed based on the expression of ribosomal proteins in several human tissues

and primary cells [26]. Group 2 proteins also show increased abundance in cancerous samples; these are involved in mRNA splicing via spliceosome biological process (based on gene ontology). Group 3 proteins likewise showed over-expression in cancer. These are composed mainly of histones present in the chromatin assembly. Histone post-translational modifications (e.g. methylation and acetylation) have been suggested to be involved in prostate tumorigenesis [27, 28]. Group 4 and Group 5 proteins show a decreased abundance in cancerous samples compared to normal tissue (Fig. 5). Some of these proteins are present in focal adhesion and participate in actin cytoskeleton organization. Protein Group 5 (highlighted in magenta in Fig. 5) is also a part of the KEGG pathways in cancer (pathway 05200). Focal adhesions are important in mediating various signals between the extracellular matrix and interacting cells and are already targeted in preclinical trials [29]. Cell adhesion molecules are crucial in tumor cell migration. Among these, we identified basal cell adhesion molecule (BCAM) protein in the current study. This protein has been shown to promote tumor cell migration by competing with integrins for binding to laminin subunit alpha-5 [30]. In the present study, the amount of both laminin subunit alpha-5 and integrin beta-1 decreased significantly in cancerous tissue, while the amount of BCAM increased in accordance with previous findings.

We performed STRING analysis also on proteins, which show a significant difference in abundance depending on cancer grade using proteins listed in Table S-3b. To select such proteins, differences between i) unified group of Grade 1, 2, and 2-3 cancer, and ii) Grade 3 cancer were considered. Like above, 15% of all proteins were selected based on a weighted t-test. In this case, the approximate limit was $p < 0.02$ and fold-change > 1.4. These represent biological changes indicative of tumor grade. STRING analysis of these proteins shows two large and one small cluster (Fig. 6). The amount of all these proteins increased in Grade 3 samples. Proteins in Group 1 are histones and present in the chromatin assembly. Group 2 proteins are mostly ribosomal

proteins participating in biological processes like translational elongation, mRNA metabolic process and belong to the KEGG ribosome pathway. Proteins of Group 3 are related to mRNA metabolic processes. In order to complement the results of the STRING analysis, we performed additional ClueGO pathway search as well. This way we determined the significance and pathway coverage of the pathways found using STRING analysis; the numerical values are summarized in Table S-4a (normal vs cancer) and Table S-4b (between cancer grades). In general, these support the conclusion of pathway analysis discussed above.

Proteins identified and quantified in the current study were compared to those published in a recent Review [10] summarizing protein biomarkers of prostate tissue. Using the described sample preparation method and limited amount of tissue sample, we observed 39 of the 69 proteins discussed (listed in Table S-5). Most of these proteins were identified by over 10 peptides, so their identifications are highly reliable. Among these 39 proteins, 10 showed similar quantitative changes as described in the Review [10], and 9 were among those used in the STRING analysis described above. The top 20 proteins showing the most significant abundance changes between healthy and cancerous subjects (based on a t-test, weighted by the fold change and protein abundance) are listed in Table 4. Among these, 18 were not listed in the Review article. Note that the results of the protein network study show that most proteins with significantly changing abundance fit into a few protein pathways. Moreover, these protein pathways (like mRNA splicing via spliceosome or chromatin assembly) are all implicated in cancer. This is an indirect evidence that protein identification and selecting those showing the most significant abundance changes are indeed reasonable, and show high probability that these might be used as prostate cancer biomarkers. This clearly demonstrates the usefulness of the described method for the identification of novel further biomarkers.

4. Conclusions

The 5-year survival rate in prostate cancer is very high (98.9%), however, it decreases significantly in the case of metastatic setting (28.5%) [31]. Thus, biomarkers that can predict therapeutic response or biological aggressiveness could further improve the therapeutic outcome of metastatic prostate cancer. High-throughput technologies such as proteomics have the potential to find new biomarkers in order to tailor therapy. A novel surface digestion protocol has been developed for the proteomic analysis of TMAs. Using the present technique, we were able to detect over 500 proteins from individual prostate TMA cores. This compares favorably with previous results on surface digestion (42 proteins on average) [17] and also with bulk digestion of scraped-off TMA cores (between 102 and 136 proteins, shown in this work). Technical reproducibility is about 3 times better than biological variability, calculated based on relative standard deviation, on correlation coefficients and on principal component analysis. The present pilot study on prostate cancer suggests that there are major proteomic differences between cancerous and healthy tissue sections which can be detected by TMA analysis. The pattern of protein abundances characterizes not only the presence of cancer but correlates with cancer grade as well. A number of proteins were identified which show significant abundance changes in prostate cancer. These show significant overlap with prostate cancer tissue biomarkers identified before [10] but include novel proteins as well. The number of proteins detected was sufficiently large to allow meaningful bioinformatics analysis. Protein abundance changes tentatively identified several protein pathways involved in prostate cancer, such as mRNA metabolic processes, mRNA splicing, chromatin assembly, focal adhesion, and the “pathways in cancer”. Several of these have been previously implicated in prostate cancer, so it is encouraging that this pilot study consisting of a relatively

small number of samples was also able to identify these. Moreover, it is promising that in this pilot study new proteins that discriminate between normal prostate tissue and prostate cancer were found. Our proteomics method may be used to supplement histological analysis, to identify novel therapeutic targets and help understand the molecular mechanisms of cancer progression. It is our hope that this method can be used to investigate prostate cancer using the remnant of prostate biopsy samples obtained from patients. A full-scale study is therefore warranted and it is in preparation based on the present results.

Supplementary material. (1) Settings of Mascot and MaxQuant software search; (2) Table S-1, MaxQuant label-free quantitative data; (3) Table S-2, List of peptides used to identify proteins reported; (4) Table S-3a, Proteins used for STRING analysis when comparing normal vs cancerous tissue; (5) Table S-3b, Proteins used for STRING analysis when comparing grade1_2_2-3 vs grade 3 tissue; (6) Table S-4a, Pathways verified by ClueGO enrichment analysis when comparing normal vs cancerous tissue; (7) Table S-4b, Pathways verified by ClueGO enrichment analysis when comparing grade1_2_2-3 vs grade 3 tissue; (8) Table S-5, List of the 39 proteins found in the present study, which were implicated in prostate cancer (Tanase et al. 2017, Oncotarget, 8, pp. 18497-18512).

Acknowledgments

LT, LD, and KV are grateful for funding from the National Research Development and Innovation Office (NKFIH PD-121187, NKFIH K-109006 and NKFIH K-119459). LT and ÁR were supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

List of Figures

Figure 1. Comparison of the number of proteins identified using surface digestion and bulk digestion of scraped-off TMA cores at low (Fig. 1A) and high (Fig. 1B) enzyme concentrations.

Each condition was analyzed in duplicates.

Figure 2. Number of proteins identified using surface digestion of Normal, Grade 2 and Grade 3 prostate cancer TMA cores in replicates.

Data in the Figure represent a compilation of the results of 36 individual experiments.

Figure 3. Non-supervised PCA analysis of label-free quantitative proteomics data obtained on 2 healthy and 10 prostate cancer TMA cores.

Cancerous samples were grouped both according to cancer stage (A) and cancer grade (B).

Figure 4. Selected examples of proteins belonging to four proteins groups based on changes between healthy and cancerous tissue: housekeeping proteins (A), decreasing proteins (B), increasing proteins (C) and special proteins (D).

Figure 5. STRING analysis of proteins with significantly changing abundance between healthy and cancerous tissue.

Only protein interactions with high confidence are shown. Protein groups highlighted in red (1-3) increase in cancer. Protein groups highlighted in blue (4-5) decrease. All proteins in Group 1 and only these belong to the KEGG ribosome pathway (red nodes). Most proteins in Group 2 belong to GO mRNA splicing via spliceosome biological process (blue nodes). Most proteins in Group 3 belong to GO chromatin assembly biological process. Most proteins in Group 4 belong to GO actin cytoskeleton organization biological process (yellow nodes). Most proteins in Group 5 (and some outside, the magenta nodes) belong to KEGG "Pathways in cancer".

Figure 6. STRING analysis of proteins significantly changing between tumor grades. All three highlighted protein groups increase in Grade 3 compared to other cancer grades.

Only protein interactions with high confidence are shown. Most proteins in Group 1 belong to GO chromatin assembly biological process (green nodes); also found in Fig. 5. Most proteins in Group 2 and Group 3 belong to GO mRNA metabolic biological process (blue nodes). Most proteins in Group 2 also belong to KEGG ribosome pathway (red nodes).

References

[1] B.W. Stewart, C.P. Wild, World Cancer Report 2014, International Agency for Research on Cancer, Lyon Cedex 08, France, 2014.

[2] L.Y. Zhao, N. Yu, T.F. Guo, Y.X. Hou, Z.Y. Zeng, X.R. Yang, P. Hu, X. Tang, J. Wang, M.R. Liu, Tissue Biomarkers for Prognosis of Prostate Cancer: A Systematic Review and Meta-analysis, Cancer Epidemiol Biomarkers Prev 23 (2014) 1047-1054.

[3] T.N. Clinton, A. Bagrodia, Y. Lotan, V. Margulis, G.V. Raj, S.L. Woldu, Tissue-based biomarkers in prostate cancer, Expert Rev Precis Med Drug Dev 2 (2017) 249-260.

[4] H. Battifora, The multitumor (sausage) tissue block - Novel method for immunohistochemical antibody testing, Lab Invest 55 (1986) 244-248.

[5] J. Kononen, L. Bubendorf, A. Kallioniemi, M. Barlund, P. Schraml, S. Leighton, J. Torhorst, M.J. Mihatsch, G. Sauter, O.P. Kallioniemi, Tissue microarrays for high-throughput molecular profiling of tumor specimens, Nat Med 4 (1998) 844-847.

[6] M. Skacel, B. Skilton, J.D. Pettay, R.R. Tubbs, Tissue microarrays: A powerful tool for high-throughput analysis of clinical specimens - A review of the method with validation data, Appl Immunohistochem Mol Morphol 10 (2002) 1-6.

[7] S. Steurer, A.S. Seddiqi, J.M. Singer, A.S. Bahar, C. Eichelberg, M. Rink, R. Dahlem, H. Huland, G. Sauter, R. Simon, S. Minner, E. Burandt, P.R. Stahl, T. Schlomm, M. Wurlitzer, H. Schluter, MALDI Imaging on Tissue Microarrays Identifies Molecular Features Associated with Renal Cell Cancer Phenotype, Anticancer Res 34 (2014) 2255-2261.

[8] A. Quaas, A.S. Bahar, K. von Loga, A.S. Seddiqi, J.M. Singer, M. Omidi, O. Kraus, M. Kwiatkowski, M. Trusch, S. Minner, E. Burandt, P. Stahl, W. Wilczak, M. Wurlitzer, R. Simon, G. Sauter, A. Marx, H. Schluter, MALDI imaging on large-scale tissue microarrays identifies molecular features associated with tumour phenotype in oesophageal cancer, Histopathology 63 (2013) 455-462.

[9] M. Galli, F. Pagni, G. De Sio, A. Smith, C. Chinello, M. Stella, V. L'Imperio, M. Manzoni, M. Garancini, D. Massimini, N. Mosele, G. Mauri, I. Zoppis, F. Magni, Proteomic profiles of thyroid tumors by mass spectrometry-imaging on tissue microarrays, Biochim. Biophys. Acta. 1865 (2017) 817-827.

[10] C.P. Tanase, E. Codrici, I.D. Popescu, S. Mihai, A.M. Enciu, L.G. Necula, A. Preda, G. Ismail, R. Albulescu, Prostate cancer proteomics: Current trends and future perspectives for biomarker discovery, Oncotarget 8 (2017) 18497-18512.

[11] E. Pin, C. Fredolini, E.F. Petricoin, The role of proteomics in prostate cancer research: Biomarker discovery and validation, Clin Biochem 46 (2013) 524-538.

[12] A. Flores-Morales, D. Iglesias-Gato, Quantitative Mass Spectrometry-Based Proteomic Profiling for Precision Medicine in Prostate Cancer, Front Oncol 7 (2017).

[13] A. Mantsiou, A. Vlahou, J. Zoidakis, Tissue proteomics studies in the investigation of prostate cancer, Expert Rev Proteomics 15 (2018) 593-611.

[14] E. Giovannucci, Y. Liu, E.A. Platz, M.J. Stampfer, W.C. Willett, Risk factors for prostate cancer incidence and progression in the health professionals follow-up study, Int J Cancer 121 (2007) 1571-1578.

[15] L. Staunton, C. Tonry, R. Lis, V. Espina, L. Liotta, R. Inzitari, M. Bowden, A. Fabre, J. O'Leary, S.P. Finn, M. Loda, S.R. Pennington, Pathology-Driven Comprehensive

518 Proteomic Profiling of the Prostate Cancer Tumor Microenvironment, Mol Cancer Res 15
519 (2017) 281-293.
520 [16] D. Iglesias-Gato, P. Wikstrom, S. Tyanova, C. Lavallee, E. Thysell, J. Carlsson, C.
521 Hagglof, J. Cox, O. Andren, P. Stattin, L. Egevad, A. Widmark, A. Bjartell, C.C. Collins,
522 A. Bergh, T. Geiger, M. Mann, A. Flores-Morales, The Proteome of Primary Prostate
523 Cancer, Eur Urol 69 (2016) 942-952.
524 [17] L. Turiak, C. Shao, L. Meng, K. Khatri, N. Leymarie, Q. Wang, H. Pantazopoulos, D.R.
525 Leon, J. Zaia, Workflow for Combined Proteomics and Glycomics Profiling from
526 Histological Tissues, Anal Chem 86 (2014) 9670-9678.
527 [18] B.C. Searle, Scaffold: a bioinformatic tool for validating MS/MS-based proteomic
528 studies, Proteomics 10 (2010) 1265-9.
529 [19] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized
530 p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat Biotech 26
531 (2008) 1367-1372.
532 [20] D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N.T.
533 Doncheva, A. Roth, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2017:
534 quality-controlled protein-protein association networks, made broadly accessible, Nucleic
535 Acids Res 45 (2017) D362-D368.
536 [21] G. Bindea, B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.H.
537 Fridman, F. Pages, Z. Trajanoski, J. Galon, ClueGO: a Cytoscape plug-in to decipher
538 functionally grouped gene ontology and pathway annotation networks, Bioinformatics 25
539 (2009) 1091-3.

[22] F.L. Bazso, O. Ozohanics, G. Schlosser, K. Ludanyi, K. Vekey, L. Drahos, Quantitative Comparison of Tandem Mass Spectra Obtained on Various Instruments, J Am Soc Mass Spectr 27 (2016) 1357-1365.

[23] E. Toth, H. Hever, O. Ozohanics, A. Telekes, K. Vekey, L. Drahos, Simple correction improving long-term reproducibility of HPLC-MS, J Mass Spectrom 50 (2015) 1130-1135.

[24] C. Arthurs, B.N. Murtaza, C. Thomson, K. Dickens, R. Henrique, H.R.H. Patel, M. Beltran, M. Millar, C. Thrasivoulou, A. Ahmed, Expression of ribosomal proteins in normal and cancerous human prostate tissue, PLoS One 12 (2017).

[25] A. Bee, Y.Q. Ke, S. Forootan, K. Lin, C. Beesley, S.E. Forrest, C.S. Foster, Ribosomal protein L19 is a prognostic marker for human prostate cancer, Clin Cancer Res 12 (2006) 2061-2065.

[26] J.C. Guimaraes, M. Zavolan, Patterns of ribosomal protein expression specify normal and malignant human cells, Genome Biol 17 (2016).

[27] Z. Chen, L.G. Wang, Q.B. Wang, W. Li, Histone modifications and chromatin organization in prostate cancer, Epigenomics 2 (2010) 551-560.

[28] D.B. Seligson, S. Horvath, T. Shi, H. Yu, S. Tze, M. Grunstein, S.K. Kurdistani, Global histone modification patterns predict risk of prostate cancer recurrence, Nature 435 (2005) 1262-1266.

[29] Y.L. Tai, L.C. Chen, T.L. Shen, Emerging Roles of Focal Adhesion Kinase in Cancer, Biomed Res Int 2015 (2015) 690690.

[30] Y. Kikkawa, T. Ogawa, R. Sudo, Y. Yamada, F. Katagiri, K. Hozumi, M. Nomizu, J.H. Miner, The Lutheran/Basal Cell Adhesion Molecule Promotes Tumor Cell Migration by

Modulating Integrin-mediated Cell Attachment to Laminin-511 Protein, J Biol Chem 288 (2013) 30990-31001.

[31] National cancer Institute, Surveillance Epidemiology and Ends Results [SEER] Program. Fast Stats; 2015 [cited June 7.2015], 2015. http://seer.cancer.gov/faststats/selections.php. (Accessed 19 December 2017).

569 **Table 1**. **Summary of diagnosis, grading and staging information of the samples analysed**

Sample	Pathology diagnosis	Grade	Stage
N1	Healthy prostate tissue	-	-
N2	Healthy prostate tissue	-	-
G1	Adenocarcinoma	Grade 1	Stage II
G2A	Adenocarcinoma	Grade 2	Stage II
G2B	Adenocarcinoma	Grade 2	Stage II
G2C	Adenocarcinoma	Grade 2	Stage IV
G2D	Adenocarcinoma	Grade 2	Stage II
G23A	Adenocarcinoma	Grade 2-3	Stage IV
G23B	Adenocarcinoma	Grade 2-3	Stage IV
G3A	Adenocarcinoma	Grade 3	Stage II
G3B	Adenocarcinoma	Grade 3	Stage III
G3C	Adenocarcinoma	Grade 3	Stage III

570

571

572

573 **Table 2. Technical reproducibility and biological variability of samples based on the label**
574 **free quantitative proteomics**

	Technical reproducibility Average	Technical reproducibility Individual	Biological variability within Stage	Biological variability within Grade	Full biological variability
Correlation coefficient	0.97*	0.95	0.92	0.94	0.85
RSD	13.8%	24.0%	30.5%	28.7%	46.7%

575 * estimated value

576

577

Table 3. Average correlation coefficients between protein fingerprints (sets of protein abundances) within and between various groups, based on cancer grade.

	Normal	**Grade 1**	**Grade 2**	**Grade 2-3**	**Grade 3**
Normal	0.98	0.90	0.88	0.81	0.68
Grade 1	0.90	---	0.94	0.89	0.84
Grade 2	0.88	0.94	0.95	0.90	0.82
Grade 2-3	0.81	0.89	0.90	0.78	0.81
Grade 3	0.68	0.84	0.82	0.81	0.90

583

584 **Table 4. Top 20 proteins showing the most significant abundance changes between healthy**
585 **and cancerous subjects (based on a t-test; weighted by the fold change and protein**
586 **abundance)**

Protein ID	Protein name	Number of identified peptides	Average MaxQuant LFQ intensity	Fold change normal vs cancerous tissue	T-test p value normal vs cancerous tissue	Described in Review [10]
SYNEM_HUMAN	Synemin	57	7.98E+06	5.8	1.76E-06	No
HBA_HUMAN	Hemoglobin subunit alpha	9	2.40E+07	3.39	2.44E-06	No
HBB_HUMAN	Hemoglobin subunit beta	12	2.47E+07	3.28	4.51E-06	No
MYH11_HUMAN	Myosin-11	132	8.76E+07	3.74	2.22E-05	No
SMTN_HUMAN	Smoothelin	17	1.79E+06	3.45	1.05E-05	No
HBD_HUMAN	Hemoglobin subunit delta	9	7.09E+05	5.01	1.5E-05	No
MYLK_HUMAN	Myosin light chain kinase. smooth muscle	21	6.45E+06	2.51	1.86E-05	No
DESM_HUMAN	Desmin	51	6.29E+07	3.39	9.22E-05	Yes
TPM1_HUMAN	Tropomyosin alpha-1 chain	24	1.10E+07	2.45	5.02E-05	Yes
LMAN1_HUMAN	Protein ERGIC-53	4	8.95E+05	0.00037	6.3E-05	No
LMOD1_HUMAN	Leiomodin-1	5	9.88E+05	2.27	3.27E-05	No
LAMA5_HUMAN	Laminin subunit alpha-5	30	1.50E+06	3.14	6.44E-05	No
PA2G4_HUMAN	Proliferation-associated protein 2G4	8	4.73E+05	0.00047	7.14E-05	No
MYL9_HUMAN	Myosin regulatory light polypeptide 9	10	8.26E+06	3.85	1.44E-04	No
FLNC_HUMAN	Filamin-C	76	1.33E+07	2.88	4.21E-04	No
RS6_HUMAN	40S ribosomal protein S6	5	1.23E+06	0.26	3.16E-04	No
RL11_HUMAN	60S ribosomal protein L11	3	8.34E+05	0.23	4.02E-04	No

FUS_HUMAN	RNA-binding protein FUS	8	1.06E+06	0.18	5.62E-04	No
FABPH_HUMAN	Fatty acid-binding protein. heart	4	5.58E+05	4.72	4.36E-04	No
MPCP_HUMAN	Phosphate carrier protein. mitochondrial	3	5.91E+05	0.0006	5.02E-04	No

587

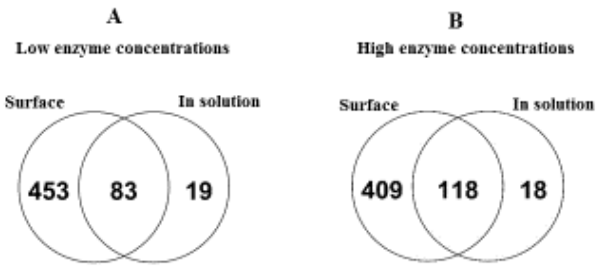
588

589

590

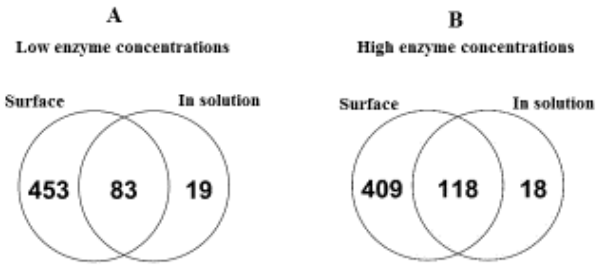
591

592



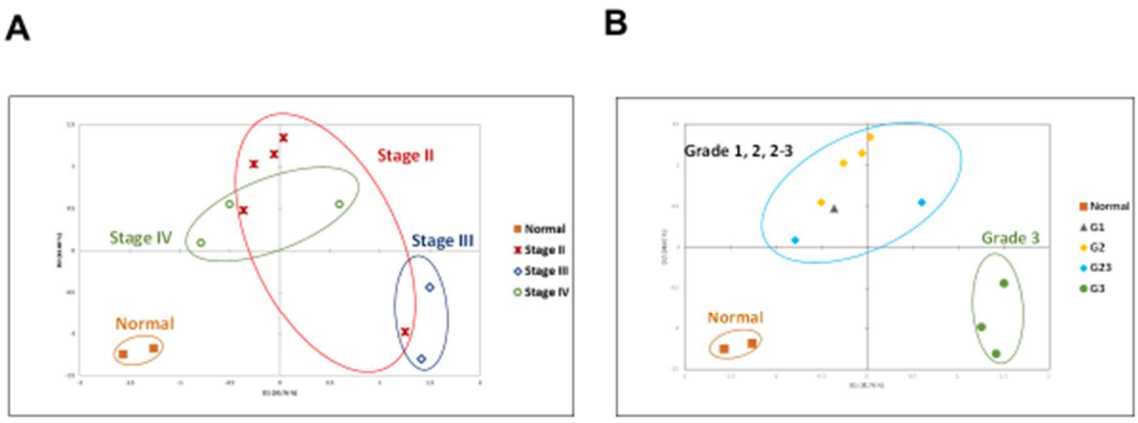
593

594 Fig 1.



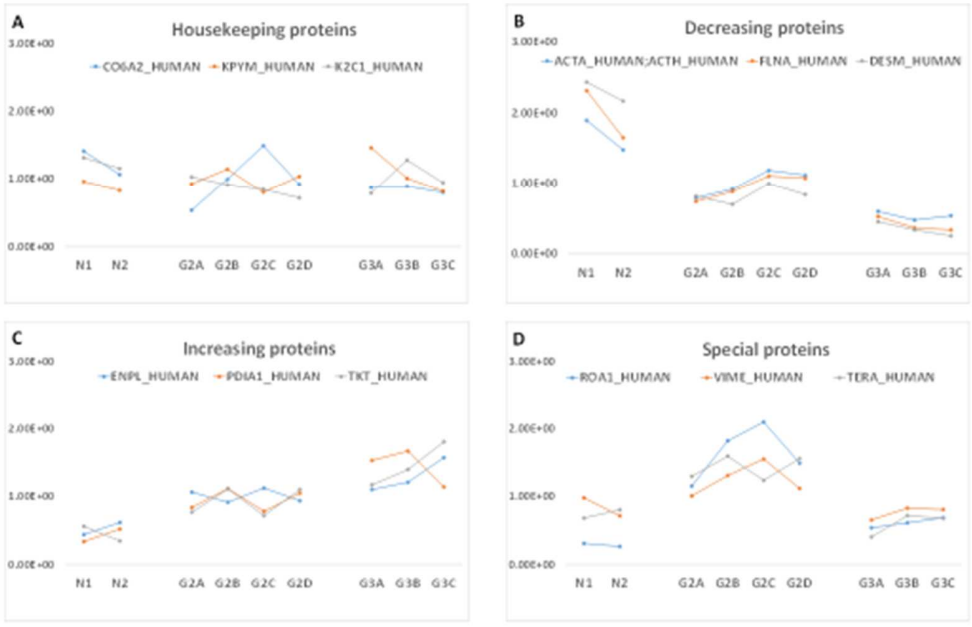
595

596 Fig 2.



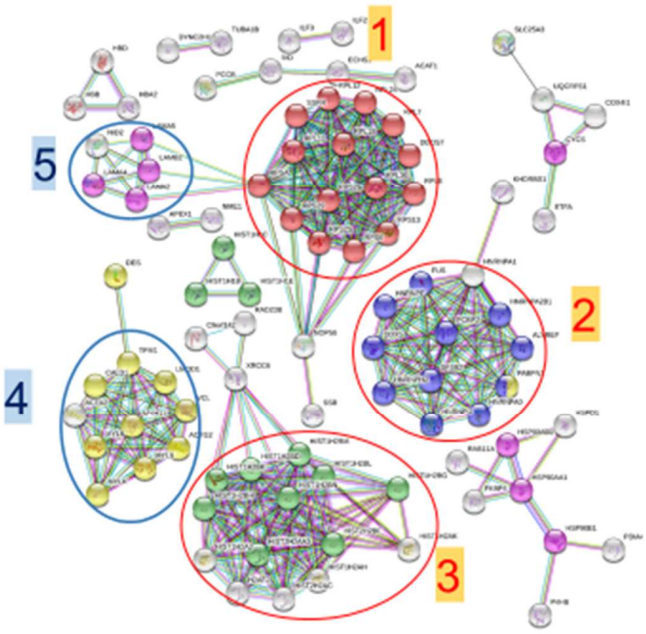
597

598 Fig 3.



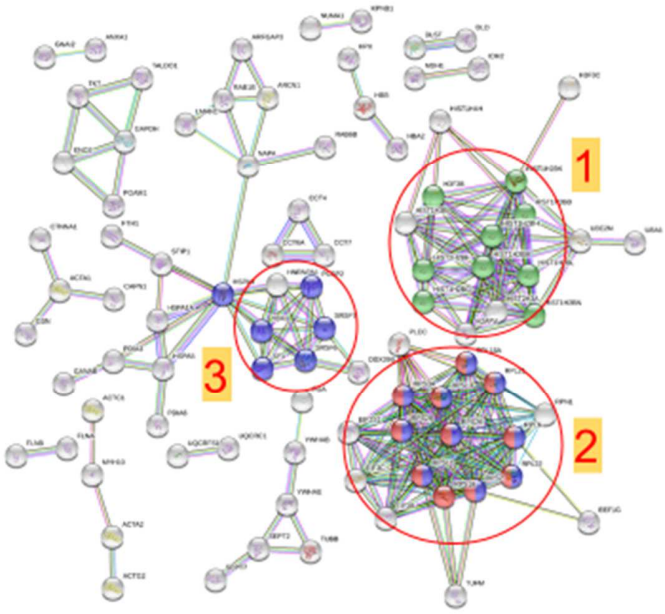
599

600 Fig 4.



601

602 Fig 5.



603

604 Fig 6.